A Note on Designs for Estimating Population Parameters

Monika Entholzner[1,2], Norbert Benda[1], Thomas Schmelter[1,2], Rainer Schwabe[1*]

[1]Otto von Guericke University, Institute for Mathematical Stochastics, PF 4120, D–39016 Magdeburg, Germany
[2]Schering AG, D–13 342 Berlin, Germany

SUMMARY

In various fields, data from repeated measurements are pooled across individuals to obtain valid estimates for population characteristics. If the individual effects are treated as random, mixed models can be fitted to the data. In the case where the same design is used for all individuals, the ordinary and the weighted least squares estimator both coincide with the average of individually fitted curves. In this situation optimal and efficient designs can be obtained.

Key words: optimal design, mixed model, weighted least squares, population parameters, repeated measurements, random regression, random coefficient regression, hierarchical model, growth curves

1. Introduction

In various fields such as pharmacokinetics, agriculture, psychology, market research, medical diagnostics, etc., there is interest in reference curves, which describe the characteristics of a whole population (e.g. Mentré et al., 1997, Retout et al., 2002, Liu et al., 2003, Curran and Hussong, 2003, Sándor and Wedel, 2002, Schwabe et al., 2001), while repeated measurements are available for individuals. Here, the design of the experimental settings, such as the time points of

* Corresponding author, e-mail: rainer.schwabe@mathematik.uni-magdeburg.de

measurements, specification of a choice set or spatial allocations, is under the control of the investigator.

Many of these relationships are of a non-linear structure. However, the quality of a design is often measured in terms of the corresponding linearization based on an asymptotic approach. Therefore, we will focus in the present note on population models specified as linear mixed models where the individual curves are described by random effects. The model will be introduced in Section 2 and the corresponding weighted least squares estimators are derived. In fact, the weighted least squares estimator is a matrix weighted mean of individually fitted curves. Such models have also been addressed as random regression, random coefficient regression, latent growth curve, latent trajectory, empirical Bayes, hierarchical or multi-level models in the literature.

In Section 3 we consider the relevant situation that all individuals are treated under the same regime. It turns out that in that case the weighted least squares and the ordinary least squares solution both coincide with the common average of the individual curves. Consequently, for the best (linear) unbiased estimator, no knowledge is required of the covariance structure for the random effects. Moreover, a simple representation of the covariance matrix is exhibited which is valid even in the case of a singular design. These models have been thoroughly investigated by Rao (1967) and many others in the literature. However, only a few attempts have been made to design such experiments. First results were obtained by Gladitz and Pilz (1982) in connection with a Bayesian approach, when individual prediction is of interest instead of estimation of the mean responses. A more general treatment has been presented by Luoma (2002) and Liski et al. (2002, p. 9, p. 52). Fedorov and Hackl (1997, p. 75) deal with the approximate theory and establish an equivalence theorem. This usually requires a large number of replications per individual. Schmelter (2005) establishes that the designs with common regime, considered here, are optimal within a more general setting. For particular models, design problems have been studied by Malyutov and Protassov (2002), Fedorov and Leonov (2004), Mentré et al. (1997), Retout et al. (2002), and Sándor and Wedel (2002) in a biomedical, pharmaceutical or market research context, respectively.

Based on the simple representation of the covariance matrix obtained, we can find optimal or, at least, efficient designs for estimating the population parameters, which is done in Section 4. The results are illustrated by some examples.

We conclude with a discussion of the results and some open problems. In particular, the results can be generalized to the comparison of treatment effects on the population parameters.

2. A Linear Mixed Model

We assume that the individual curves are governed by a common functional structure and that there are additional observational errors. Observations are given by a common linear model

$$Y_{ij} = \mathbf{f}(x_{ij})^{\mathrm{T}} \mathbf{b}_i + \varepsilon_{ij} ,$$

$j = 1,\ldots,m_i$, and $i = 1,\ldots,n$. Here, Y_{ij} is the jth observation for individual i at the experimental setting x_{ij} subject to a random error ε_{ij}, m_i is the number of observations for individual i, and n is the total number of individuals. $\mathbf{f} = (f_1,\ldots,f_p)$ are known regression functions and $\mathbf{b}_i = (b_{i1},\ldots,b_{ip})^{\mathrm{T}}$ is the p-dimensional vector of parameters for the individual curve associated with subject i. For example, all curves may be straight lines (individual linear regression) or all curves are polynomials of a fixed degree (individual quadratic regression etc.). However, the experimental settings x_{ij} need not be merely restricted to the time points of measurements ($x_{ij} = t_{ij}$, $t_{i1} < \ldots < t_{im_i}$) as in growth curve models (see Christensen, 2001, p. 49), but may also describe additional factors such as treatments, measurement methods or even dosage or spatial characteristics, to cover all possible applications in mind.

The individual parameters $\mathbf{b}_i$ are assumed to be random, drawn from a homogeneous population, with mean $E(\mathbf{b}_i) = \beta$ and covariance matrix $\mathrm{cov}(\mathbf{b}_i) = \sigma^2 \mathbf{D}$. Moreover, the individual effects are assumed to be uncorrelated to each other, $\mathrm{cov}(\mathbf{b}_i, \mathbf{b}_{i'}) = \mathbf{0}$ for $i \neq i'$, and to the error terms, $\mathrm{cov}(\mathbf{b}_i, \varepsilon_{i'j}) = \mathbf{0}$. Finally, homoscedastic errors are assumed,

$$\mathrm{Var}(\varepsilon_{ij}) = \sigma^2 > 0, \mathrm{cov}(\varepsilon_{ij}, \varepsilon_{i'j'}) = 0 \text{ for } (i,j) \neq (i',j') .$$

Under normality assumptions this means that $\mathbf{b}_i \sim N(\beta, \sigma^2 \mathbf{D})$, $\varepsilon_{ij} \sim N(0, \sigma^2)$ and all individual effects and observational errors are independent.

To allow for fixed effects across the population, the dispersion matrix $\mathbf{D}$ may be singular.

In the present note we are mainly interested in the vector of population parameters $\beta = (\beta_1,\ldots,\beta_p)^{\mathrm{T}}$, which characterizes the mean response of the whole population.

Accordingly, the model equation can be rewritten as a standard linear mixed model

$$Y_{ij} = \mathbf{f}(x_{ij})^{\mathrm{T}} \beta + \mathbf{f}(x_{ij})^{\mathrm{T}} (\mathbf{b}_i - \beta) + \varepsilon_{ij} ,$$

where, now, $\mathbf{f}(x_{ij})^{\mathrm{T}} \beta$ is the mean response, $E(Y_{ij}) = \mathbf{f}(x_{ij})^{\mathrm{T}} \beta$, and the random component $\mathbf{f}(x_{ij})^{\top}(\mathbf{b}_i - \beta) + \varepsilon_{ij}$ has zero mean and variance $\sigma^2(1 + \mathbf{f}(x_{ij})^{\top}\mathbf{D}\mathbf{f}(x_{ij}))$. Observations belonging to the same individual are correlated, $\mathrm{cov}(Y_{ij}, Y_{ij'}) = \sigma^2\mathbf{f}(x_{ij})^{\top}\mathbf{D}\mathbf{f}(x_{ij'}), j \neq j'$, while observations coming from different individuals are not, $\mathrm{cov}(Y_{ij}, Y_{i'j'}) = 0, i \neq i'$. This is exactly the setting of a completely randomized design for the individuals.

Now, for each individual i, denote by $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{im_i})^{\top}$ and $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{im_i})^{\top}$ the vectors of corresponding observations and observational errors, respectively, and by $\mathbf{F}_i = (\mathbf{f}(x_{i1}), \ldots, \mathbf{f}(x_{im_i}))^{\top}$ the associated $m_i \times p$ design matrix. Then

$$\mathbf{Y}_i = \mathbf{F}_i\beta + \mathbf{F}_i(\mathbf{b}_i - \beta) + \varepsilon_i$$

describes the contribution of individual i with mean response $E(\mathbf{Y}_i) = \mathbf{F}_i\beta$ and covariance $\mathrm{cov}(\mathbf{Y}_i) = \sigma^2\mathbf{V}_i$, where

$$\mathbf{V}_i = \mathbf{I}_{m_i} + \mathbf{F}_i\mathbf{D}\mathbf{F}_i^{\top}$$

and $\mathbf{I}_m$ denotes the $m \times m$ identity matrix. For the whole study population of n individuals we combine the individual observation and error vectors $\mathbf{Y}_i$ and ε_i to form the N-dimensional population vectors

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_n \end{pmatrix} \quad \text{and} \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix},$$

respectively. Here, $N = \sum_{i=1}^{n} m_i$ is the total number of observations across the whole sample of individuals. Accordingly, the $N \times p$ population design matrix becomes

$$\mathbf{F} = \begin{pmatrix} \mathbf{F}_1 \\ \vdots \\ \mathbf{F}_n \end{pmatrix}.$$

The whole observation vector $\mathbf{Y}$ can, then, be written as

$$\mathbf{Y} = \mathbf{F}\beta + \mathbf{G}(\mathbf{b} - \mathbf{1}_n \otimes \beta) + \varepsilon,$$

where $\mathbf{G}$ is block diagonal,

$$\mathbf{G}=\begin{pmatrix}\mathbf{F}_1 & & 0\\ & \ddots & \\ 0 & & \mathbf{F}_n\end{pmatrix}, \quad \text{and} \quad \mathbf{b}=\begin{pmatrix}\mathbf{b}_1\\ \vdots \\ \mathbf{b}_n\end{pmatrix}$$

is the (random) vector of individual parameters. Note that $E(\mathbf{b})=\mathbf{1}_n\otimes\beta$ where $\mathbf{1}_n$ is an n-dimensional vector with all entries equal to one, and "$\otimes$" denotes the common Kronecker product of vectors and matrices, respectively. Consequently, the expectation of $\mathbf{Y}$ equals $\mathbf{F}\beta$ and the covariance matrix $\operatorname{cov}(\mathbf{Y})=\sigma^2\mathbf{V}$ is block diagnonal,

$$\mathbf{V}=\begin{pmatrix}\mathbf{V}_1 & & 0\\ & \ddots & \\ 0 & & \mathbf{V}_n\end{pmatrix},$$

where the diagonal blocks are the covariance matrices for the single individuals. This structure is due to the fact that both the individual parameter vectors $\mathbf{b}_i$ and the individual errors ε_i are uncorrelated,

$$\operatorname{cov}(\mathbf{b})=\sigma^2\mathbf{I}_n\otimes\mathbf{D},\ \operatorname{cov}(\varepsilon)=\sigma^2\mathbf{I}_N \text{ and } \operatorname{cov}(\mathbf{b},\varepsilon)=\mathbf{0}.$$

We first consider the regular case that the design matrix $\mathbf{F}$ has full column rank p. If the dispersion matrix $\mathbf{D}$ is known, the best linear unbiased estimator for the population parameter β is given by the weighted least squares estimator

$$\hat{\beta}_{WLS}=\left(\mathbf{F}^{\top}\mathbf{V}^{-1}\mathbf{F}\right)^{-1}\mathbf{F}^{\top}\mathbf{V}^{-1}\mathbf{Y}$$

according to the well-known Gauss-Markov theorem.

Due to the block diagonal structure of $\mathbf{V}$, the weighted least squares estimator may be rewritten as

$$\begin{aligned}\hat{\beta}_{WLS}&=\left(\sum_{i=1}^{n}\mathbf{F}_i^{\top}\mathbf{V}_i^{-1}\mathbf{F}_i\right)^{-1}\sum_{i=1}^{n}\mathbf{F}_i^{\top}\mathbf{V}_i^{-1}\mathbf{Y}_i\\ &=\left(\sum_{i=1}^{n}\mathbf{F}_i^{\top}\mathbf{V}_i^{-1}\mathbf{F}_i\right)^{-1}\sum_{i=1}^{n}\mathbf{F}_i^{\top}\mathbf{V}_i^{-1}\mathbf{F}_i\hat{\mathbf{b}}_i,\end{aligned}$$

where $\hat{\mathbf{b}}_i$ is any (weighted) least squares solution for the individual curve

$$\hat{\mathbf{b}}_i=\left(\mathbf{F}_i^{\top}\mathbf{V}_i^{-1}\mathbf{F}_i\right)^{-}\mathbf{F}_i^{\top}\mathbf{V}_i^{-1}\mathbf{Y}_i$$

by means of the normal equations. Here, $\mathbf{A}^{-}$ denotes any arbitrary generalized inverse of $\mathbf{A}$, i. e. $\mathbf{A}\,\mathbf{A}^{-}\mathbf{A}=\mathbf{A}$.

As $\mathbf{V}_i \mathbf{F}_i = (\mathbf{I}_{m_i} + \mathbf{F}_i \mathbf{D} \mathbf{F}_i^\top)\mathbf{F}_i = \mathbf{F}_i(\mathbf{I}_p + \mathbf{D}\mathbf{F}_i^\top \mathbf{F}_i) = \mathbf{F}_i \mathbf{U}_i$ for some $p \times p$ matrix $\mathbf{U}_i$, the individual weighted least squares solution coincides with the ordinary least squares solution

$$\hat{\mathbf{b}}_i = (\mathbf{F}_i^\top \mathbf{F}_i)^- \mathbf{F}_i^\top \mathbf{Y}_i$$

(see Zyskind, 1967). The individual estimators $\hat{\mathbf{b}}_i$ therefore do not require knowledge of the dispersion matrix $\mathbf{D}$. This is in accordance with the fact that random coefficients are immaterial when observations are available for a single individual only.

Note that the individually fitted parameters $\hat{\mathbf{b}}_i$ are not the best linear unbiased predictors as they ignore the covariance structure $\mathbf{D}$ and the information that can be derived from the other individuals in the population. In fact, the best linear unbiased predictors are weighted combinations of the individually fitted parameters and the population estimates (see e. g.McCulloch,2003,p.13).

If all individual design matrices $\mathbf{F}_i$ have full column rank p, then individual curves can be fitted uniquely. The (unique) weighted least squares estimator $\hat{\beta}_{WLS}$ is a matrix weighted mean of the individually estimated parameters. However, in this averaging procedure the dispersion matrix $\mathbf{D}$ is involved.

The covariance matrix of the weighted least squares estimator is given by

$$\mathrm{cov}(\hat{\beta}) = \sigma^2 \left(\sum_{i=1}^{n} \mathbf{F}_i^\top \mathbf{V}_i^{-1} \mathbf{F}_i \right)^{-1} .$$

For an early review of the analysis in the regular case when $\mathbf{F}$ has full rank we refer to Spjøtvoll (1977).

In the case where $\mathbf{F}$ does not have full column rank p, it is still true that

$$\begin{aligned} \hat{\beta}_{WLS} &= (\mathbf{F}^\top \mathbf{V}^{-1} \mathbf{F})^- \mathbf{F}^\top \mathbf{V}^{-1} \mathbf{Y} \\ &= \left(\sum_{i=1}^{n} \mathbf{F}_i^\top \mathbf{V}_i^{-1} \mathbf{F}_i \right)^- \sum_{i=1}^{n} \mathbf{F}_i^\top \mathbf{V}_i^{-1} \mathbf{F}_i \hat{\mathbf{b}}_i \end{aligned}$$

is a weighted least squares solution, whatever generalized inverse of $\mathbf{F}^\top \mathbf{V}^{-1} \mathbf{F}$ $= \sum_{i=1}^{n} \mathbf{F}_i^\top \mathbf{V}_i^{-1} \mathbf{F}_i$ and individual least squares solutions $\hat{\mathbf{b}}_1, \ldots, \hat{\mathbf{b}}_n$ are chosen. Moreover, if the linear aspect $\psi = \psi(\beta) = \mathbf{L}\beta$ is identifiable, i.e. $\mathbf{L} = \mathbf{K}\mathbf{F}$ for a suitable matrix $\mathbf{K}$, then $\psi = \mathbf{L}\hat{\beta}_{WLS}$ is the best linear unbiased estimator of ψ with covariance matrix

$$\operatorname{cov}(\hat{\psi}) = \sigma^2 \mathbf{L} \left(\sum_{i=1}^{n} \mathbf{F}_i^{\mathrm{T}} \mathbf{V}_i^{-1} \mathbf{F}_i \right)^{-} \mathbf{L}^{\mathrm{T}}$$

which depends, in general, on the dispersion $\mathbf{D}$.

In many applications the choice of the experimental settings x_{ij}, $j = 1, \ldots, m_i$, $i = 1, \ldots, n$, is under the control of the investigator and may be chosen from a design region χ of possible settings. As the quality of an estimator measured in terms of its covariance matrix heavily depends on the settings x_{ij}, the design of these settings is an important task. Moreover, the (co)variances depend on the number n of individuals and on the numbers m_i of repeated measurements for each individual. Even if we assume (as we do here) that these sample sizes n and m_i are fixed, there are usually still many possibilities for individual designs $(x_{i1}, \ldots, x_{im_i})$. Therefore, in what follows, we will restrict our attention to the practically relevant situation in which all individuals have the same design.

3. Designs with Common Regime

In practical situations such as human or animal pharmacokinetic studies or medical diagnostics there are often external restrictions, e.g. technical implementations (see Figure 1 for a spatial scheme in medical diagnostics), which force the experiment to be performed with identical regimes for all individuals. This means that for each individual the number m_i of repeated measurements equals m, and the experimental settings $x_{ij} = x_j$ are identical across all individuals.

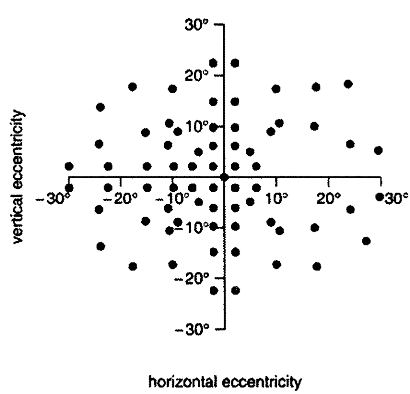

Figure 1. TCC-grid (Schwabe et al., 2001) for the perimetric investigation of the visual ability (differential luminance sensitivity, right eye, blind spot omitted)

Thus both the design matrices $\mathbf{F}_i = \mathbf{F}_1$ and the associated covariance matrices $\mathbf{V}_i = \mathbf{V}_1$ coincide for all individuals (see Liski et al., 2002, p. 9, Fedorov and Hackl, 1997, p. 75, for an extension to approximate designs, and Schmelter, 2005, for the fact that such designs turn out to be optimal within a more general setting). If all individuals have the same design, the weighted least squares estimator simplifies to the mean of the individually fitted parameters

$$\widehat{\beta}_{WLS} = \left(n\mathbf{F}_1^\top \mathbf{V}_1^{-1} \mathbf{F}_1\right)^{-1} \sum_{i=1}^{n} \mathbf{F}_1^\top \mathbf{V}_1^{-1} \mathbf{Y}_i$$
$$= \frac{1}{n} \sum_{i=1}^{n} \hat{\mathbf{b}}_i$$

if $\mathbf{F}_1$ has full column rank p.

Hence, calculation of the weighted least squares estimator $\widehat{\beta}_{WLS}$ does not require knowledge of the dispersion matrix $\mathbf{D}$. Moreover, $\widehat{\beta}_{WLS}$ coincides with the ordinary least squares estimator

$$\widehat{\beta}_{WLS} = \widehat{\beta}_{OLS} = \left(n\mathbf{F}_1^\top \mathbf{F}_1\right)^{-1} \sum_{i=1}^{n} \mathbf{F}_1^\top \mathbf{Y}_i$$

(see e.g. Rao, 1967, or Bischoff, 1992, for a related result). In general, for every linear identifiable aspect $\psi = \psi(\beta) = \mathbf{L}\beta$ the matrix $\mathbf{L}$ can be represented as $\mathbf{L} = \mathbf{K}_1 \mathbf{F}_1$, and the weighted least squares estimator, which is the best linear unbiased estimator, equals

$$\widehat{\psi} = \frac{1}{n} \sum_{i=1}^{n} \widehat{\psi}_i,$$

where

$$\widehat{\psi}_i = \mathbf{L}\left(\mathbf{F}_1^\top \mathbf{F}_1\right)^{-} \mathbf{F}_1^\top \mathbf{Y}_i$$

is the corresponding individual estimate. In particular, for estimating the mean response $\mu(x) = \mathbf{f}(x)^\top \beta$ at a prespecified experimental setting x, we obtain

$$\hat{\mu}(x) = \frac{1}{n} \sum_{i=1}^{n} \hat{\mu}_i(x),$$

where $\hat{\mu}_i(x) = \mathbf{f}(x)^\top (\mathbf{F}_1^\top \mathbf{F}_1)^{-} \mathbf{F}_1^\top \mathbf{Y}_i$ is the fitted value of the individual curve evaluated at x. This fact is depicted in Figure 2 for data obtained from twelve healthy subjects aged thirty to forty where the design of Figure 1 was used (Schwabe et al., 2001).

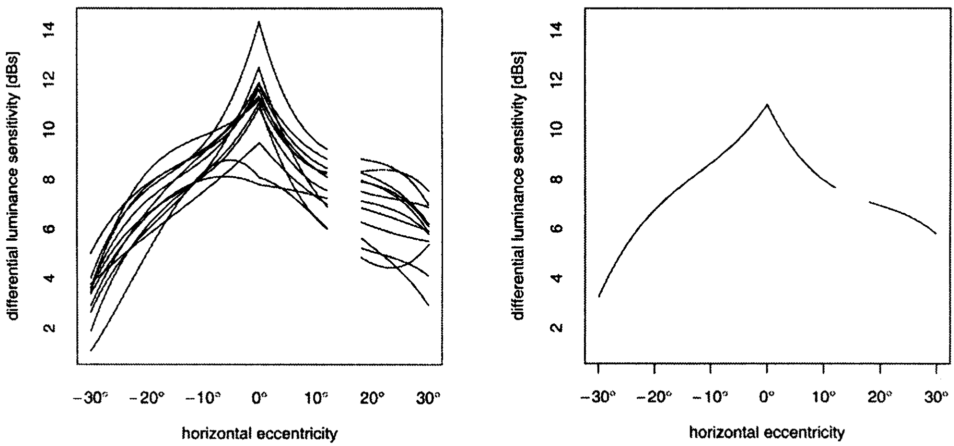

Figure 2: individual profiles (left panel) and estimated population mean (right panel) for the visual ability (right eye, horizontal cross section at vertical eccentricity 0 °, blind spot omitted)

For every identifiable aspect $\psi = \psi(\beta) = \mathbf{L}\beta$ with $\mathbf{L} = \mathbf{K}_1 \mathbf{F}_1$ the covariance matrix of the (weighted) least squares estimator is given by

$$
\begin{aligned}
\operatorname{cov}(\hat{\psi}) &= \frac{\sigma^2}{n} \mathbf{L}\left(\mathbf{F}_1^\top \mathbf{V}_1^{-1} \mathbf{F}_1\right)^- \mathbf{L}^\top \\
&= \frac{\sigma^2}{n} \mathbf{L}\left(\left(\mathbf{F}_1^\top \mathbf{F}_1\right)^- + \mathbf{D}\right)\mathbf{L}^\top .
\end{aligned}
$$

This can easily be seen from the fact that the ordinary and weighted least squares estimators coincide,

$$
\begin{aligned}
\operatorname{cov}(\hat{\psi}) &= \operatorname{cov}(\psi(\hat{\beta}_{OLS})) \\
&= \frac{\sigma^2}{n} \mathbf{L}\left(\mathbf{F}_1^\top \mathbf{F}_1\right)^- \mathbf{F}_1^\top \mathbf{V}_1 \mathbf{F}_1 \left(\mathbf{F}_1^\top \mathbf{F}_1\right)^- \mathbf{L}^\top \\
&= \frac{\sigma^2}{n} \mathbf{K}_1 \mathbf{F}_1\left(\mathbf{F}_1^\top \mathbf{F}_1\right)^- \mathbf{F}_1^\top \left(\mathbf{I}_m + \mathbf{F}_1 \mathbf{D} \mathbf{F}_1^\top\right)\mathbf{F}_1\left(\mathbf{F}_1^\top \mathbf{F}_1\right)^- \mathbf{F}_1^\top \mathbf{K}_1^\top \\
&= \frac{\sigma^2}{n} \mathbf{K}_1 \mathbf{F}_1\left(\left(\mathbf{F}_1^\top \mathbf{F}_1\right)^- + \mathbf{D}\right)\mathbf{F}_1^\top \mathbf{K}_1^\top \\
&= \frac{\sigma^2}{n} \mathbf{L}\left(\left(\mathbf{F}_1^\top \mathbf{F}_1\right)^- + \mathbf{D}\right)\mathbf{L}^\top .
\end{aligned}
$$

By letting $\mathbf{L} = \mathbf{F}_1^{\mathrm{T}} \mathbf{V}^{-1} \mathbf{F}_1$ we see, as a by-product, that $(\mathbf{F}_1^{\mathrm{T}} \mathbf{F}_1)^{-} + \mathbf{D}$ is a generalized inverse of $\mathbf{F}_1^{\mathrm{T}} \mathbf{V}_1^{-1} \mathbf{F}_1$.

Note that the covariance matrix decomposes additively into the covariance matrix $\mathbf{L}(\mathbf{F}_1^{\mathrm{T}} \mathbf{F}_1)^{-} \mathbf{L}^{\mathrm{T}}$ of the fixed effects model, $Y_{ij} = \mathbf{f}(x_{ij})^{\mathrm{T}} \beta + \varepsilon_{ij}$ $(\mathbf{D} = \mathbf{0})$, neglecting the random effects, and the contribution $\operatorname{cov}(\psi(\mathbf{b}_i)) = \mathbf{L}\mathbf{D}\mathbf{L}^{\mathrm{T}}$ from the original dispersion of the random effects. If $\mathbf{F}_1$ is of full column rank p, then the covariance matrix for the whole parameter vector decomposes to $\operatorname{cov}(\hat{\beta}_{WLS}) = \frac{\sigma^2}{n}((\mathbf{F}_1^{\mathrm{T}} \mathbf{F}_1)^{-1} + \mathbf{D})$ (see Rao, 1967, Mentré et al., 1997, Retout et al., 2002, and McCulloch, 2003, for particular cases).

Though the best linear unbiased estimators do not depend on the dispersion matrix $\mathbf{D}$, it is evident that their covariance matrices are strongly influenced by the variability of the random effects.

It is worth mentioning that, instead of the above cross-sectional consideration, the estimation can also be performed in a longitudinal way based on the averaged observations $\overline{\mathbf{Y}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{Y}_i$,

$$\hat{\beta}_{WLS} = (\mathbf{F}_1^{\mathrm{T}} \mathbf{F}_1)^{-1} \mathbf{F}_1^{\mathrm{T}} \overline{\mathbf{Y}}$$

(see Carrig et al., 2004).

4. Optimal Designs

The design problem at hand is to determine $x_1, \ldots, x_m$ such that

$$\operatorname{cov}(\hat{\beta}_{WLS}) = \frac{\sigma^2}{n}\left(\left(\mathbf{F}_1^{\mathrm{T}} \mathbf{F}_1\right)^{-1} + \mathbf{D}\right)$$

or

$$\operatorname{cov}(\hat{\psi}) = \frac{\sigma^2}{n}\left(\mathbf{L}\left(\mathbf{F}_1^{\mathrm{T}} \mathbf{F}_1\right)^{-}\mathbf{L}^{\mathrm{T}} + \mathbf{L}\mathbf{D}\mathbf{L}^{\mathrm{T}}\right),$$

respectively, becomes minimal.

If the linear aspect ψ is one-dimensional, $\psi = \mathbf{c}^{\mathrm{T}} \beta$ for some p-dimensional vector $\mathbf{c}$, we have a complete ordering of the estimators as $\operatorname{cov}(\hat{\psi})$ reduces to the variance $\operatorname{Var}(\hat{\psi})$ which is a real number. However, in general we have to compare matrices which are only partially ordered.

Therefore, it becomes necessary to consider real-valued functionals of the corresponding covariance matrices, such as the trace which measures the expected Euclidean distance of the estimator from the true parameter value, or the determinant (sometimes called "generalized variance") which describes the volume of the confidence ellipsoid for the parameter vector.

We first restrict our attention to linear criteria which aim at minimizing the trace

$$\text{trace}\left(\mathbf{L}\left(\left(\mathbf{F}_1^\top \mathbf{F}_1\right)^- + \mathbf{D}\right)\mathbf{L}^\top\right) = \text{trace}\left(\mathbf{L}\left(\mathbf{F}_1^\top \mathbf{F}_1\right)^- \mathbf{L}^\top\right) + \text{trace}\left(\mathbf{L}\mathbf{D}\mathbf{L}^\top\right)$$

of the normalized covariance matrix $\text{cov}(\hat{\psi})$ for $\hat{\psi} = \mathbf{L}\hat{\beta}$. This covers the A-criterion, $\mathbf{L} = \mathbf{I}_p$, of minimizing the expected Euclidean distance $E(\|\hat{\beta} - \beta\|^2)$, the IMSE-criterion, $\mathbf{L}^\top\mathbf{L} = \int \mathbf{f}(x)\mathbf{f}(x)^\top \lambda(dx)$, of minimizing the integrated mean squared error for the estimated response, and all one-dimensional c-criteria for a particular linear combination $\psi(\beta) = \sum_{j=1}^{p} c_j \beta_j$, $\mathbf{L} = \mathbf{c}^\top = (c_1, \ldots, c_p)$, which aim at minimizing the variance of the one-dimensional least squares estimator $\mathbf{c}^\top \hat{\beta}$ of $\mathbf{c}^\top \beta$.

Due to the additive decomposition of the covariance matrix, all linear criteria coincide with their fixed effects model counterpart,

$$\text{trace}\left(\mathbf{L}\left(\mathbf{F}_1^\top \mathbf{F}_1\right)^- \mathbf{L}^\top\right)$$

up to an additive constant, $\text{trace}(\mathbf{L}\mathbf{D}\mathbf{L}^\top)$. Hence, for a linear criterion function every optimal design in the reduced fixed effects model

$$Y_{ij} = \mathbf{f}(x_j)^\top \beta + \varepsilon_{ij},$$

neglecting the individual effects ($\mathbf{D} = \mathbf{0}$), which minimizes $\text{trace}(\mathbf{L}(\mathbf{F}_1^\top\mathbf{F}_1)^- \mathbf{L}^\top)$, turns out to be optimal in the corresponding mixed effects model.

For every design $(x_1, \ldots, x_m)$, characterized by its design matrix $\mathbf{F}_1$, the efficiency in the mixed model is given by

$$\text{eff}_\mathbf{L} = \frac{\text{trace}\left(\mathbf{L}(\mathbf{F}_*^\top \mathbf{F}_*)^- \mathbf{L}^\top\right) + \text{trace}\left(\mathbf{L}\mathbf{D}\mathbf{L}^\top\right)}{\text{trace}\left(\mathbf{L}(\mathbf{F}_1^\top \mathbf{F}_1)^- \mathbf{L}^\top\right) + \text{trace}\left(\mathbf{L}\mathbf{D}\mathbf{L}^\top\right)},$$

where $\mathbf{F}_*$ is the design matrix of an optimal design. It is readily seen that, for every design, the efficiency $\text{eff}_\mathbf{L}$ in the mixed model is at least as large as the efficiency in the corresponding fixed effects model without individual effects $(\mathbf{D} = \mathbf{0})$.

Consequently, a highly efficient design in the fixed effects model also performs very well in the mixed model. Note that, in general, the efficiency specifies the amount of information obtained by using $\mathbf{F}_1$ instead of the optimal design matrix $\mathbf{F}_*$. More precisely, $n \cdot \mathrm{eff} \leq n$ is the approximate number of individuals needed if the optimal $\mathbf{F}_*$ is used to obtain the same criterion value as for the design represented by $\mathbf{F}_1$ applied to n individiuals.

Other criteria do not share the property of linearity, and hence the optimal design in the mixed effects model may differ substantially from its fixed model counterpart. For example, for the commonly used D-criterion, the determinant $\det\left((\mathbf{F}_1^\top \mathbf{F}_1)^{-1} + \mathbf{D}\right)$ of the standardized covariance matrix has to be minimized, which essentially describes the volume of a confidence ellipsoid for β. Obviously, minimization of $\det\left((\mathbf{F}_1^\top \mathbf{F}_1)^{-1} + \mathbf{D}\right)$ and $\det\left((\mathbf{F}_1^\top \mathbf{F}_1)^{-1}\right)$ may yield different solutions. Examples are given below.

Similar findings hold for the G-criterion which aims at minimizing the worst value of the variance function

$$\sup_{x\in\chi} \mathbf{f}(x)^\top \left((\mathbf{F}_1^\top \mathbf{F}_1)^{-1} + \mathbf{D}\right)\mathbf{f}(x)$$

of prediction over the design region χ.

The efficiencies for the D- and G-criterion are

$$\mathrm{eff}_D = \left(\frac{\det((\mathbf{F}_*^\top \mathbf{F}_*)^{-1}+\mathbf{D})}{\det((\mathbf{F}_1^\top \mathbf{F}_1)^{-1}+\mathbf{D})}\right)^{1/p} \text{ and } \mathrm{eff}_G = \frac{\sup_{x\in\chi} \mathbf{f}(x)^\top ((\mathbf{F}_*^\top \mathbf{F}_*)^{-1}+\mathbf{D})\mathbf{f}(x)}{\sup_{x\in\chi} \mathbf{f}(x)^\top ((\mathbf{F}_1^\top \mathbf{F}_1)^{-1}+\mathbf{D})\mathbf{f}(x)}, \text{ respectively.}$$

Example 1. We consider the most simple design setting of a dichotomous design variable $x \in \{1,2\}$. Here, $x = 1$ and $x = 2$ describe two different methods of measurement of certain characteristics for each individual. Measurements may be replicated and are subject to measurement errors. The individual effects b_{i1} and b_{i2} are assumed to be random with means μ_1 and μ_2, respectively, for the two measurement methods. Furthermore, we suppose that the number m of replications is the same for all individuals, e.g. due to constraints on the number of blood samples that can be taken. Finally all individuals should be treated in the same way due to organizational constraints, i. e. for each individual $m(1)$ observations are made with method 1 while $m(2) = m - m(1)$ observations are made with method 2.

With a standard dummy coding, $\mathbf{f}(1) = (1,0)^\top$ and $\mathbf{f}(2) = (0,1)^\top$, we obtain for the standardized covariance matrix

$$(\mathbf{F}_1^\top \mathbf{F}_1)^{-1} + \mathbf{D} = \begin{pmatrix} m(1)^{-1} + d_{11} & d_{12} \\ d_{12} & m(2)^{-1} + d_{22} \end{pmatrix},$$

where d_{ij} are the corresponding entries in the initial dispersion matrix $\mathbf{D}$ of the random effects b_{i1} and b_{i2}, $\mathrm{Var}(b_{ij}) = \sigma^2 d_{jj}$, $\mathrm{cov}(b_{i1}, b_{i2}) = \sigma^2 d_{12}$.

Hence, for the linear A- and IMSE-criterion the optimal design assigns $m(1) = m(2) = m/2$ observation to both methods if m is even ($m(1) = m(2) \pm 1 = (m \pm 1)/2$ if m is odd). This is the equal allocation rule, which is optimal if the random effects are ignored.

For the D-criterion we have to minimize $(d_{11} + 1/m(1))(d_{22} + 1/m(2))$. The D-optimal values of $m(1)$ are presented in Table 1 for certain combinations of d_{11}, d_{22}, and m.

Table 1. D-optimal values of $m(1)$ for $d_{11} = 1$, $d_{22} = 1, 2, 5, 10, 100$, and $m = 2, 4, 10, 100$

	d_{22}				
m	1	2	5	10	100
2	1	1	1	1	1
4	2	2	3	3	3
10	5	6	7	7	9
100	50	59	69	76	91

Note that $1 \le m(1) \le m - 1$ in order to achieve identifiability. If m is large the proportion $\alpha = m(1)/m$ may be considered to vary continuously between 0 and 1. By standard calculus the optimal proportion has to satisfy $d_{11}\alpha^2 - d_{22}(1-\alpha)^2 = (1-2\alpha)/m$. For m large this is achieved for $m(j)$ proportional to $d_{jj}^{-1/2}$, i. e. m(1) $\approx \sqrt{d_{22}}/(\sqrt{d_{11}} + \sqrt{d_{22}})$. The entries in Table 1 corresponding to $m = 100$ are already close to the limiting values. In this situation the D-efficiency of the equal allocation rule, $m(1) \approx m(2)$, may drop to 70%.

Similarly, for the G-criterion, we have to minimize $\max(d_{11} + 1/m(1), d_{22} + 1/m(2))$. For m large and d_{11}, d_{22} small (e. g. $d_{11} = 2/m, d_{22} = 0$) the efficiency of the equal allocation rule may be reduced to 85%.

Typically, in most applications, the number m of replications will be small. Of course, for saturated designs with $m = 2$, the equal allocation rule, $m(1) = m(2) = 1$, is the only design which identifies the parameters and is therefore optimal. For $m = 4$, $d_{11} = 2/3$, $d_{22} = 0$ the equal allocation rule is substantially outperformed by the design $m(1) = 3$ and $m(2) = 1$ for the G-criterion (efficiency 86%) and by the contrary design $m(1) = 1$ and $m(2) = 3$ for the D-criterion (efficiency 97%). Note that the G-criterion is more sensitive to random effects. In contrast to the case of fixed effects models, the D- and G-criteria lead to opposite optimal designs, a counter-intuitive fact, which, as far as we know, has not yet been addressed in the literature.

Example 2. For the linear regression model $Y_{ij} = b_{i0} + b_{i1}\, x_i + \varepsilon_{ij}$ the optimal design points will in many cases be located at the endpoints when the design region χ is an interval, by virtue of majorization arguments (see e. g. Krafft, 1978, p. 250). At least, such designs will be very efficient. As the D- and G-criterion are not affected by reparameterizations, if the dispersion matrix **D** is transformed accordingly, the linear regression setup restricted to the optimal endpoints can be identified with the dichotomous situation of Example 1 and the corresponding findings carry over. In particular, for the saturated case, m = 2, it can easily be seen that it is opimal to take one obervation at each of the two endpoints of the underlying interval. Further results on optimal designs in a linear or quadratic random regression setup are given by Liski et al. (2002, p. 52).

Example 3. Consider the multiple linear regression, $\mathrm{Y}_{ij} = b_{i0} + b_{i1}\, x_{j1} + b_{i2}\, x_{j2} + \varepsilon_{ij}$, $\mathbf{x}_j = (x_{j1}, x_{j2})$, on the unit square. In the fixed effects model ($\mathbf{D} = \mathbf{0}$) a D- and G-optimal design can be obtained as the cross-product of its one-dimensional counterparts (see Schwabe, 1996, p. 52). In the present random effects setting this general result does not hold true unless severe restrictions are made on the structure of the dispersion matrix $\mathbf{D}$. However, if interest is solely in the direct effects $(\beta_1, \beta_2) = (E(b_{i1}), E(b_{i2}))$ and the corresponding individual effects are uncorrelated, $\operatorname{cov}(b_{i1}, b_{i2}) = 0$, then the cross-product design remains D-optimal for (β_1, β_2) by virtue of a majorization argument. This result can be extended to general additive models as treated in Schwabe (1996, section 5.1). In this situation, full as well as fractional factorial designs remain optimal.

Example 4. In population pharmacokinetics exponential decay models are considered. As the functional relation is non-linear in the parameter, in general, $Y_{ij} = \eta(x_j, \mathbf{b}_i) + \varepsilon_{ij}$, only locally optimal designs can be derived which are based on the asymptotic covariance matrix obtained by linearization (see e.g. Mentréetal, 1997, or Retoutetal, 2002).

For example, in the one-parameter exponential decay model, $\eta(x,b) = e^{-bx}$, $0 \le x < \infty$, the asymptotic information obtained by linearization is given by

$$\left(\frac{\partial \eta(x,\beta)}{\partial \beta} \right)^2 = x^2\, e^{-2\beta x}$$

which attains its maximum at the optimal design point $x^* = 1/\beta$. As there is only one parameter, all criteria coincide and $x^* = 1/\beta$ is also optimal in the random effects model.

In the two-parameter exponential decay model, $\eta(x,\mathbf{b}) = b_1\, e^{-b_2 x}$, $0 \le x < \infty$, the asymptotic information is obtained from the linearized model

$$Y_{ij} \approx \mathbf{f}_{\beta}(x_j)^{\mathrm{T}} \beta + \varepsilon_{ij} ,$$

where

$$\mathbf{f}_{\beta}(x_j)^{\mathrm{T}} = \left(\frac{\partial \eta(x,\beta)}{\partial \beta_1}, \frac{\partial \eta(x,\beta)}{\partial \beta_2} \right) = (e^{-\beta_2 x}, -\beta_1 x e^{-\beta_2 x})$$

is the vector of partial derivatives of the mean response η with respect to the parameters β_1 and β_2. In the fixed effects model the D-optimal design points are given by 0 and $x^* = 1/\beta_2$, with equal weights. In the random effects model the saturated D-optimal design points, $m = 2$, are 0 and x_0, where x_0 depends on both the dispersion matrix $\mathbf{D}$ and β_2. In most situations x_0 is slightly larger than x^*. However, the efficiency of the design $(0, x^*)$ remains surprisingly high.

5. Discussion

Population parameters require a particular design strategy which may differ substantially from the corresponding fixed effects model ignoring the individual random effects. If identical designs are used across individuals, we can average over individually fitted curves and ignore the dispersion when estimating. Moreover, the covariance matrix, and hence the design problem, becomes simpler. For linear criteria, the optimal fixed effects designs turn out to be optimal also for the population parameters in a random effects environment, while for other criteria certain adjustments become necessary.

The results can be extended to treatment comparisons and similar situations: Also, in the situation where two or more groups are to be compared, the weighted least squares estimators for the single treatment curves coincide with the corresponding averages of the individual curves within each treatment group, if within each group all individuals receive the same (group-specific) design. Optimal designs can be obtained by optimizing within each treatment group, separately, according to the group specific dispersion. In particular, if the dispersion structure is identical across all groups, an optimal solution can be obtained where across all groups the individuals are treated under the same design. Such a procedure is advantageous, if not necessary, in the setting of a blind or double-blind study, where the design is not allowed to contain any information which might reveal the particular treatment to the individual (or the study nurse). Moreover, in this situation the variances for single parameters coincide across the groups, which justifies the commonly used two-stage approach to perform a one-way analysis of variance on the individually obtained parameters. Similarly, it can be seen that the concept of identical individual designs also proves to be optimal for more than one treatment factor, in the presence of block effects, and in cross-over settings.

Topics of further research include designs that may differ across individuals, the case of an unknown dispersion matrix at the design stage, if the criterion is not linear, and optimal designs for the variance components and for predictions, respectively. For example, consideration of designs that differ across individuals becomes essential when fewer observations are available per individual than the number of parameters, so that no individual curve can be fitted.

Acknowledgement. The authors are grateful to Ulrike Graßhoff for helpful discussions and to two anonymous referees for valuable comments.

REFERENCES

Bischoff W. (1992): On exact D-optimal designs for regression models with correlated observations. *Annals of the Institute of Statistical Mathematics*, 229–238.

Carrig M. M., Wirth R. J. and Curran P. J. (2004): A SAS macro for estimating and visualizing individual growth curves. *Structural Equation Modeling*, 132–149.

Christensen R. (2001): *Advanced Linear Modeling, Second Edition.* Springer, New York.

Curran P. J. and Hussong A. M. (2003): The use of latent trajectory models in psychopathology research. *Journal of Abnormal Psychology*, 526–544.

Fedorov V. V. and Hackl P. (1997): *Model-Oriented Design of Experiments. Lecture Notes in Statistics*. Springer, New York.

Fedorov V. V. and Leonov S. (2004): Optimal designs for regression models with forced measurements at baseline. In *mODa 7 - Advances in Model-Oriented Design and Analysis* (A. Di Bucchianico, H. Läuter and H. P. Wynn, eds.). Physica, Heidelberg, 61–69.

Gladitz J. and Pilz J. (1982): Construction of optimal designs in random coefficient regression models. *Mathematische Operationsforschung und Statistik, Series Statistics*, 371–385.

Krafft O. (1978): *Lineare Statistische Modelle und optimale Versuchspläne.* Vandenhoeck & Ruprecht, Göttingen.

Liski E. P., Mandal N. K., Shah K. R. and Sinha B. K. (2002): *Topics in Optimal Design. Lecture Notes in Statistics*. Springer, New York.

Liu Z., Reinhardt F., Bünger A. and Reents R. (2004): Derivation and calculation of approximate reliabilities and daughter yield-deviations of a random regression test-day model for genetic evaluation of dairy cattle. *Journal of Dairy Science*, 1896–1907.

Luoma A. (2000): *Optimal Designs in Linear Regression Models. Acta Universitatis Tamperensis*. Academic Dissertation, University of Tampere, Department of Mathematics, Statistics and Philosophy.

Malyutov M. B. and Protassov R. S. (2002): LAN, LAM and convergence of iterative ML estimates; optimal design in Gaussian one-way mixed model. *Journal of Statistical Planning and Inference* 249–268.

McCulloch C. E. (2003): *Generalized Linear Mixed Models. NSF-CBMS Regional Conference Series in Probability and Statistics* . Institute of Mathematical Statistics, Beachwood.

Mentré F., Mallet A. and Baccar D. (1997): Optimal design in random-effects regression models. *Biometrika* , 419–442.

Rao C. R. (1967): Least squares theory using an estimated dispersion matrix and its applications to measurement of signals. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* . University of California Press, 355–372.

Retout S., Mentré F. and Bruno R. (2002): Fisher information matrix for non-linear mixed-effects models: evaluation and application for optimal design of enoxaparin population pharmacokinetics. *Statistics in Medicine* , 2623–2639.

Sándor Z. and Wedel M. (2002): Profile construction in experimental choice designs for mixed logit models. *Marketing Science* , 455–475.

Schmelter T. (2005): On the optimality of single-group designs in linear mixed models. Otto-von-Guericke-Universität Magdeburg, Fakultät für Mathematik, Preprint Nr. 02/2005.

Schwabe R. (1996): *Optimum Designs for Multi-Factor Models. Lecture Notes in Statistics* . Springer, New York.

Schwabe R., Vonthein R., Ata N., Pätzold J., Dietrich T. J. and Schiefer U. (2001): Modeling the hill of vision. In *Perimetry Update 2000/2001* (M. Wall and R. P. Mills, eds.). Kugler, The Hague, 71–79.

Spjøtvoll E. (1977): Random coefficient regression models. A review. *Mathematische Operationsforschung und Statistik, Series Statistics* , 69–93.

Zyskind G. (1967): On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. *Annals of Mathematical Statistics* , 1092–1109.